Darrell A. Poirier
006-110-300

# System for transferring personalized matter from one computer to another.

INVENTOR: Darrell A. Poirier, **590 Prospect Street, Woodstock, CT 06281**

## References to Related Applications

This application is a continuation of Provisional Patent Application No. 60/156,638, filed September 29, 1999, and Provisional Patent Application No. 60/214,504, filed on June 28, 2000.

## STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

This invention has been created without the sponsorship or funding of any federally sponsored research or development program.

## BACKGROUND OF THE INVENTION

1. Field of the Invention

This invention relates to voice recognition. It explains methodologies for performance, reliability, and accuracy improvement and a standardized way of measurement. The patent also introduces machine independent user mobility between different voice recognition systems. It addresses a method for enabling large vocabulary voice recognition for masses of people without needing training. It describes how to apply the technology to a new style of interactive real time voice to text feedback for a handheld transcriber to replace the previous non-interactive handheld transcribers. And, finally it explains how the technology can be used to enable voice mail to text using large vocabulary voice recognition engines. Some applications that this invention applies to include; speech recognition, appliances (stationary and handheld), robotics, voice mail, voice command, and noise recognition.

2. Description of the Prior Art

In general terms, large vocabulary voice recognition products that are currently in the market follow the clone PC market strategy. Clone PC components are integrated using minimum requirements delineated by the voice recognition software applications. Therefore on average, the state of the art prior to this invention is buying a personal



EXPRESS MAIL NO. EL522965955US

computer that is designed as a general purpose computing device and using that as a Large Vocabulary Voice Recognition (LVVR) system. While this approach is typically used throughout the computer industry today, it often leaves LVVR users frustrated with accuracy and performance. This is especially true when applying the technology to hand held transcriber type of devices like a tape recorder or voice recorder to solid state memory devices.

Another problem is machine dependency. LVVR systems require training to enable the system to understand the person doing the speaking. The upfront training prior to an acceptable level of accuracy can vary from hours to upwards of weeks as the system learns a specific user. This large investment of time and effort is a per machine cost virtually causing a user to be machine dependent.

When trying to sort out accuracy and performance a user can spend much effort, time and money trying to determine what are the best options for the cost. This has led to frustration and many dollars being wasted with the result being that LVVR applications are bought and left sitting on the shelf or discarded. Additional results include the voice recognition industry products being criticized as inadequate, insufficient, and a bad investment in general. From a user and investors point of view, the LVVR voice recognition industry has been tarnished.

While in some ways the LVVR industry is in recovery, the basic problems of **standard performance** and **standard accuracy, machine dependency, speaker dependency, mobility,** and **method of estimating cost** is still typical. This patent targets these specific problems.

## 3. BRIEF SUMMARY OF THE INVENTION

In summary this invention defines, labels, and solves the problems of inadequate performance and accuracy by developing the measurement of Reliable Accuracy Performance (RAP rate). The invention also solves the problem of being machine dependent without loosing a quality RAP rate. The accomplishments are achieved by improving measurement techniques (q), standard level of components (s), and functionality suited for voice recognition (i), combined with new concepts based in software to enable new functionality into the industry.

This invention contains concepts and descriptions including: Specification and integration of hardware component functionality and features for voice recognition. The ability to create Transportable Voice Models called Voice Model Mobility (VMM) that enables machine independent voice recognition. Using Super Voice Model(s) (SVM) which allows speaker independent voice recognition. Metering techniques that can measure RAP rate enabling a set of features to be measured and qualified to a minimum specific level for LVVR systems. The ability to build a service industry archiving and supplying voice models to users on an as needed basis by using the internet, email, US

mail on various transfer mediums like disk, network, credit card strips, etc.. This invention addresses the problems listed below helping users and the voice recognition industry to move "up a level" and forward another step to the future.

- Performance measured in Speech To Text Delay being unpredictable
- Insufficient performance and features for using voice recognition
- Incompatibilities of hardware and software
- Training required for each voice recognition machine a person uses
- Inconsistent accuracy when using multiple voice recognition machines
- Outdated or stale Voice Models that misrepresent the person speaking
- Lack of speaker independent ability for large vocabulary voice recognition
- The lack of large vocabulary voice recognition for handheld computers
- Time and effort lost due to trial and error creation of LVVR systems

4. DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

**Invention descriptions**

**LX. Voice Model Mobility (VMM)**

***Brief History***

When using Large Vocabulary Voice Recognition (LVVR) applications, training of the voice recognition software is involved to allow the software to understand the uniqueness of a specific user. This uniqueness is stored in a file, operating system parameter list (like registry entries) and other formats that defines specific parameters (Voice Model) for a specific user. The ability to unplug these parameters and data from one machine, put them on some transfer medium like disk, optical disk, floppy, network disk drive, etc. and move them to another machine is defined by this patent as Voice Model Mobility or VMM.

Voice Model Mobility (VMM) was originally conceived due to the problem of having to train multiple LVVR machines for a single person's voice. This was discovered when experimenting with LVVR applications. It was determined that a better way to use multiple machines was to separate the files and parameters that characterize the user, package the files and parameters as a voice model and move them to a transferable medium for installation into another system.

Voice models can and should be independent of voice recognition applications allowing the user of a voice model to plug into and use any voice recognition application. Voice Model Mobility version 1.0 (VMM V1.0) is the first step toward the concept of modular plug-able voice models. This concept enables new features to be incorporated into voice

models to provide enhancements on a wide variety of applications (e.g. security, learning aid for people speaking, singing, and language, language translation, voice mail to text, games, guidance, etc.).

Prior to VMM, voice models did not exist and there was no easy way to move these specific user parameters and data between machines. Several experiments were done in effort of understanding why applications did not support such features. From these experiments it was discovered that the lack of ability to create and move a voice model was not technical.

The first experiment was to use the backup and save feature provided with the Dragon Professional voice recognition application. The problems encountered when trying to accomplish this included a different filename when restoring the user from when the user was saved. Another problem was the limitation of where the backup could be saved. In other words the voice model was not mobile.
The second experiment was to copy the voice model files directly to another location and then copy them back to use them. In some cases this approach appeared to work although it took some trial and error until the exact files that needed to be copied were discovered. Crashes and hangs occurred often. Problems encountered prior to successful file copies included; user voice files contamination, the system hanging when trying to open a specific user, or the Dragon application no longer finding the user for opening. Although this approach sometimes yielded success it was discovered that the user would have to be created first, and then the files could be copied. This was due to registry entries were not setup as part of the copy process. A Visual Basic prototype was coded using this method for user interface experimentation.

The third effort included investigation of the system registry to determine if Dragon was setting any parameters using the registry. This was found to be true and solved the final problems. The current version of VMM is coded in the C programming language.

*Method of fixing the problem*

A "Voice Model" is defined in this patent as a signal, information, or electronic data file that is information and/or a parameter representation of a person's voice or a noise. A Voice Model contains attributes that characterize items such as formants, phonemes, speaking rate, pause length, etc. for a given user. One use for a voice model that contains data and parameters of a specific user is that it allows the user to take advantage of Large Vocabulary Voice Recognition (LVVR) applications. LVVR applications require such parameters to be known to achieve accurate voice recognition. All approaches to LVVR (e.g. Acoustic phonetic, Pattern recognition, Artificial intelligence, Neural networks, etc.) require training. Training is required to create a reference pattern from which decisions are made using templates or statistical models (e.g. Markov Models and Hidden Markov Models) as to the probability of the word to be translated.

Any noise that can be captured contains within it parameters that have the ability to enable creation of a voice model. VMM allows a voice model to be unplugged from the Dragon Professional Voice Recognition application and be moved to a transferable medium. The medium and voice model can then be moved and plugged into another system running the Dragon Professional Voice Recognition application.

The next step is translation of the users voice model to be recognized by other LVVR applications (e.g. IBM ViaVoice, L&H, and Philips Speech Pro). An alternative to translation of voice models is a standard voice model format. In overview the mechanics of translating voice models between LVVR applications include; 1) An information file is built identifying which parameters are needed for each LVVR. 2) The parameters are read from one LVVR and translated to an LVVR common file format (.lvr). 3) Based on the known parameters, selection is made for the remaining parameter components based from knowledge of other voice models with similar parameters. 5) The components are added and linked in the .lvr file. The file is then translated to the desired voce model format to create the final voice model. 6) The voice model is plugged-in to the destination LVVR using the VMM techniques.

The next section is the help file from the VMM prototype. This prototype has been tested, debugged, and upgraded and is now being used by many people. This prototype happen to use CD write media for the transfer medium. The current version works with high capacity floppy disk, network drives, CD media, internet network drives, etc..

Given future data compression and larger media capacities, the goal would be to put Voice Models on credit card type magnetic strips requiring personalized identification to enable the models similar to credit cards and ATM cards of today.

### *Theory of operation*

While the concept of Voice Model Mobility could be applied with any voice recognition software, application or hardware, the VMM software application developed and used here for example purposes is layered upon Dragon Systems Voice Recognition Professional Application. The transfer medium is CD read/write disk, but could just as well have been floppy disk, network, ROM, credit card strip, or other means of storing data.

(SEE FIGURE 1)

**<u>Sequence of events: User clicks "Move voice model to CD-Writer" button:</u>**

1. A folder labeled "Users" is created on the destination media.
2. A "users.ini" file is created in the Users folder. This file is a logical translation from Username to a user file name that Dragon will open.
3. VMM then creates and writes the user specific registry information into a file called

VMMinfo.txt

4. A "user" specific folder will be created in the Users folder. There are several files in the user specific folder.
5. The user as a result of the Dragon training process creates the files listed below. These files are copied to the CDWriter using the standard Dragon directory structure. Files included are:

audioin.dat
Current Folder containing:
    topics (configuration file)
    options (configuration file)
    global.DVC
    Voice folder
        DD10User.sig
        DD10User.usr
    GeneralE Folder
        dd10voc1.voc
        dd10voc2.voc
        dd10voc3.voc
        dd10voc4.voc
        General.voc
    Shared Folder
        archive.voc

6. These files and related registry parameters information make up the voice model for this example for the Dragon Professional application.

**<u>When the user clicks move voice model to hard drive the following events take place:</u>**

1. VMM pops up a window asking which drive contains the voice model.

(SEE FIGURE 2)

2. After the drive is selected, VMM looks on the selected drive for the Users folder containing the voice models, specifically, the users.ini file.
3. If VMM does not find any users, a window pops up saying that no users were found with an OK button to click returning the user to the previous screen.

(SEE FIGURE 3)

4. VMM then asks the user to select one of the voice models it found on the selected drive.

(SEE FIGURE 4)

5. VMM then reads the file VMMInfo.txt file to determine the appropriate registry settings.
6. If the user already exists, VMM will prompt the user to ask if the user files should be overwritten.
7. If the user responds by clicking the OK button, then steps 7 onward will be executed, other wise VMM will go back to the main VMM screen.
8. If there is no user specific folder, then VMM creates the user specific folder in the standard Dragon directory structure otherwise it uses the existing folder.
9. VMM then copies all the specific user files listed above to the specific user folder.
10. VMM then sets the up the registry for the selected user.

This is one approach that we have chosen to implement Voice Model Mobility. We have discussed other implementations including translating voice models to a unique file formats to include additional information, or encryption for security reasons.

## #$ Voice Model Mobility Help

Help Topics



Voice Model Mobility (VMM) can be used to transport copies of your "voice model" on CD between National Voice systems. This allows many users the ability to use many different systems without having to train each system to recognize each voice. Also, every time you write your voice model to CD you have created a backup.



(SEE FIGURE 1)

### # What is Voice Model Mobility (VMM) ?

A voice model is the parameters that allow a computer to recognized a voice of an individual person or noise. Usually this refers to a digitized model of a speech pattern that is used for voice recognition. VMM is the ability to move the Voice

---

# HIDD_HELPAPP_DIALOG

$ Table of Contents

# HID_DEFINE_VMM

Model between computers allowing each computer to use the Voice Model with a standard level of performance on each system. Without VMM each system needs to be trained and configured separately.

(SEE FIGURE 5)

**# Copying the Voice Model to CD**

To copy your voice model to CD-ROM you need to place a writeable CD or CD/RW disk into the CD-Writer. If this is a new CD and has not been used to move a voice model before, the VMM utility will prepare the CD prior to writing the voice model.

To move a voice model press the button "Move voice model to CD writer".

CD writers on National Voice systems are always located at the D:\ drive

(SEE FIGURE 6)

**# Copying the Voice Model to the Hard Disk**

To copy your voice model to the hard disk insert the CD-ROM containing the voice model into the CD drive. Press the button "Move voice model to hard drive.

Select the name of the voice model and click "Ok". Coping the voice model takes about 2 minutes.

---

## # System Requirements

Version 1.0 of VMM supports Dragon Professional editions using Windows 95 and Windows 98. The following versions of Dragon are supported:

Version 3.0
Version 3.52
Version 4.0

VMM requires a high capacity media type (e.g. CD/R, CD/RW, high capacity floppy (100 MB+), zip drives, hard disk, network mapped drives and internet mapped drives).

## LXI. Super Voice Models (SVM)

### *Brief History*

Machine and speaker dependant machines describe the LVVR voice recognition systems of today. When using Large Vocabulary Voice Recognition (LVVR) applications two problems are experienced. These are machine dependency and speaker dependency. A user of LVVR needs to train a voice recognition machine and is virtually tied to the machine (machine dependent). If another voice recognition system is used, that system needs to be trained as well. If many people want to use any non-specific voice recognition system for LVVR (speaker independent), it is not possible with the technology of today with the exception that each person would have to train each machine to be used which is usually not feasible.

### *Method for solving problem*

Super Voice Models have the ability to achieve speaker independent voice recognition. The technology that enables this ability is VMM. Given that Voice Models will be available for transfer, they will be collected. By having a collection of voice models available for analysis and modification, a new type of voice model can be created, derived from the parameters available from the collection. In general the Super Voice Model involves information and processes with the final output result a real time Voice Model to recognize the person speaking at the time. The new Voice Model can optionally be calculated real time or prior for a given person.

As we look toward the future, computer voice recognition is moving toward noise recognition using one of many methods (e.g. training, pre-programming, learned through

the experience of artificial intelligence or expert systems). If this technology is applied to the analysis of plane crashes, as an example, it could help to an understanding failures leading back to root causes.

*Theory of operation*

The SVM concept is based upon having information about Voice Models organized, readily available, and ready to execute to statistically and accurately synthesize a Voice Model thread that can be used for any given speaker at the time. The overview process flow is as follow:

VMM-> Voice Models-> Collection-> Analysis-> Modifications-> Voice Model Thread

The many voice models collected are categorized defining group parameters. As a person starts speaking to the machine the real-time voice is measured and categorized selecting the proper group parameters applying the group parameters to that individual person. Based on the known group parameters, selections of the unknown parameters are made from statistical a model.

**LXII. RAP metering (Recognition Accuracy Performance)**

*Brief History*

Many professional people use more than 1 computer to accomplish their daily task. When more than 1 computer is used for voice recognition, accuracy and performance may not be consistent. This was discovered through experimentation with voice recognition packages and was verified when talking with doctors and lawyers and other professional people that use speech recognition. These users described accuracy for example, at an estimated 94 percent but all claimed that they didn't know accurately what the accuracy was. Other statements made included how accuracy would vary when using an assortment of machines for LVVR.

During experimentation with voice recognition applications the same problems were discovered. It was decided that indicators would be needed to accurately measure accuracy over word input count. The metrics that mattered to people were two. These included; reliable accuracy, delay time of spoken word until text is displayed in an application on the screen. These two measurements are key in what people expect from a quality voice recognition system. This patent labels them as Reliable Accuracy Performance Rate or "RAP Rate".

*Methods of fixing problems*

Accuracy and performance are two important metrics that determine the adequacy of a Large Vocabulary Voice Recognition (LVVR) system. When using multiple LVVR systems, the expectation is to get at least the same RAP Rate between the systems. To accomplish a consistent RAP rate an accurate measure must first be referenced. Then, based on the RAP metrics, decisions can be made as to the adequacy. If improvement is the goal then, logical decisions can be made as to what area to investigate improving based on the results from the RAP meter.

The RAP meter measures accuracy by having a person read text available on the LVVR system. As a person reads the text the words are translated into text. The RAP meter compares the original text with the text that was translated using the LVVR system and responds back to the user with an accuracy measurement in percentage of correct translated words (e.g. 96 % correct). Thus, Accuracy % = words incorrect / words correct. Incorrect words are highlighted for display to the user.

For performance, the RAP meter records the time (Tstart) that sound was input and subtract from the time (Tend) that text is displayed on the screen for editing in a text application. Thus, Performance = Tend - Tstart

The RAP rate meter concept is a feature of the Voice Model Mobility concept providing an indication of successful operation of moving a voice model. The RAP meter can also be provided as a separate application for certification of LVVR applications. As an example, companies that advertise RAP certification could be charged a per package fee to create revenues.

*RAP Rate meter theory of operation*

Reliable Accuracy Performance Rate (RAP Rate) is defined by this patent as percentage of accuracy delivered with measured delay time from word spoken to visible text in a document. Components that effect RAP rate include hardware components, software components, and a person or "user". From this, the following can be stated.

**Reliable Accuracy Performance Rate = User + System + Quality of components + Integration**

**Or**

**RAP rate = u + s + q + i**

Where a user "u" is defined as the person speaking to a voice recognition system. The system "s" is defined as a system trained to recognize a person's voice. Quality of components "q" is defined as the hardware and software component functionality that is appropriate for LVVR, and finally integration "i" defined as how the components are

combined together including the merging of hardware, software, and parameters focusing on providing optimal voice recognition.

For example, if a system has a reliable accuracy of 96% and a reliable performance of 1 second, then the RAP would equal 96% at 1second or a RAP rate of 96 to1.
A large vocabulary voice recognition system today including quality components and good integration can deliver a RAP rate of approximately 96% at 4 seconds (96 to 4).

The RAP Rate equation components can be further defined:

*(Affecting reliable performance)*

***Quality "q"***

Defined as a compatibility of components and functionality that are well-matched for LVVR.

$$q = \text{CPU margin \%} \,/\, (\text{\% of app in memory} \,/\, (\text{\% of app in memory} - (\text{KB Cache} \,/\, \text{cache hit rate})/60) - (\text{A/D conversion time} + \text{bus throughput latency})$$

The equation above indicates aspects of hardware that can be changed to achieve an improved RAP rate focusing on the metric of *Performance*. The performance result is measured in time (seconds for current technology). Performance can be a positive or negative value. A positive value indicates delay while a negative value is additional resources that can be used for other task. The "delay" in the performance definition will never be zero. It may not be perceivable to a user, but it can be measured by using the RAP meter. The items listed in the *quality components* equation directly affects the Performance = Tend – Tstart equation.

*(Affecting reliable accuracy)*

***Integration "i"***

$$i = \text{System parameters} + \text{Application parameters} - \text{Incompatibilities} - \text{Other task executed} - \text{Throughput resistance}$$

The integration aspect of RAP rate affects reliable accuracy. System parameters include

hardware (i.e. microphones, sound AD conversion devices, etc.) and software, (i.e. firmware, bios, operating systems, applications/utilities and parameters). Computer parameters are designed to accomplish many different jobs and as a result contain parameters that can conflict with a specific goal such as LVVR. Setting up software parameters to ensure the capabilities for LVVR are enabled at all levels is needed. Integration ties directly to RAP in the measure of reliable accuracy.

*Mechanics of implementing the Performance function of a RAP Rate meter*

The components used to get real time capture and performance measurement includes; Application handles to indicate applications loaded and used for LVVR. Audio input device IRQ and I/O address range and software driver IO function calls to indicate when the A/D translation has started. Speech recognition function calls (e.g. RealTimeGet and TimeOutGet) to indicate when the voice recognition engine has started and completed the translation. Video board IRQ and I/O address range and software driver IO function calls to determine when the text is being displayed to the editor on the screen. As words are spoken into a microphone, trigger points are set to indicate when each section of the process has completed its task. The following steps indicate how the RAP meter functions;

1) (Setup) Application used for LVVR is identified

2) (Tstart) A/D time is measured by logging the time the driver gets sound input. This can be accomplished through a peek message or for MSWindows ;InmChannelAudio::IsIncoming, HRESULT IsIncoming(void);

3) (Pstart) Determine and log the when the speech processing engine has received the sound by using a function call (i.e. RealTimeGet).

4) (Pend) Determine and log the time when the speech engine has completed the translation using a function call (i.e. TimeOutGet).

5) (Tend) Determine when the graphics driver has displayed the text using a peek message or for MSWindows a function call (i.e. UI Text Event; TEXT_VALUE_CHANGED).

6) (Report) Calculate the times. For general performance Tend – Tstart will supply the performance delay. For further resolution to determine areas of throughput resistance, steps 2 and 3 can be used.

**LXIII. Specified standardized hardware model for voice recognition**

***Brief History***

From 3 years of investigation and experimentation with various hardware, software applications and operating systems it was discovered that a broad range of accuracy and performance is realized. The experimentation was done using hardware and software components from the following companies:

| CPU's | MB Chipsets | Sound devices | Applications | Environments |
|---|---|---|---|---|
| Intel | Intel | Creative | Dragon | MSWindows 95 |
| Cyrix | VIA | Telex | L&H | MSWindows 98 |
| AMD | | Yamaha | Kerswell | MSWindows NT 4.0 |
| | | Acer | Microsoft | |
| | | Labtec | IBM | |
| | | Kensiko | | |
| | | Corel | | |
| | | Shure | | |
| | | Parrot | | |
| | | Sony | | |
| | | Radio Shack | | |

From this work it was discovered that specific hardware features could enhance accuracy and performance (measured in speech to text delay) while other features or the lack of resulted in reduced accuracy and performance. While the LVVR products advise for minimum requirements, features needed for optimal LVVR is not stated.

### *Methods of fixing problems*

Using RAP Rate as an indicator, voice recognition system components can be qualified for configurations optimal for LVVR. It was discovered that component features could be identified, documented and put into a process to provide a standard level of features and functionality to address RAP rate.

These specific features include:

<u>Optimal features to enhance RAP Rate</u>

- High-speed microprocessors
- Robust floating point features
- Large on chip and off chip cache memory 512 kb or more
- High-capacity/fast main memory (optimal 512 megabytes)
  - Sound input device with performance focused on input in the range of the human voice
  - An operating system specifically configured (tuned) for the application of voice recognition including:
  - Removing any throughput resistance including processes that require main CPU

clock cycles but don't provide advantage to LVVR.

- Removing operating system resources that use main memory or run in the background like schedulers, virus checking, or utilities that execute polling at specific time intervals or triggers.
- Removing applications that use main CPU floating point and moving that work to other microprocessors.
- Ensuring that any operating system or applications being used return allocated memory back to being available and not left locked out by the LVVR application.

These specific features are not included recommendations with the "off the shelf" voice recognition application packages from vendors like IBM, Dragon, or L&H. The features are dedicated to the task of voice recognition and can be packaged as such to create a large vocabulary voice recognition appliance.
Originally, to measure and characterize the hardware for large vocabulary voice recognition experimentation, tools and indicators that were readily available in the industry and part of the operating systems were used. Using these tools measurements could be acquired and a determination was made as to hardware resources needed. Then a manual process of measurement was used in effort to refine what further hardware parameters would be best. From this work an automated test methodology was built to allow production mode for development and manufacturing to be put in place to characterize the hardware faster.

### *Theory of operation*

This invention is a process that measures specific hardware features necessary to support optimal Large Vocabulary Voice Recognition (LVVR) Reliable Accuracy Performance (RAP) Rate. During the measurement values are inserted into process sheets allowing controlled steps to be followed. Using this technique processes can be developed for a production mode of LVVR system development and manufacture.

For development the methods include a hardware components selection process based on investigation of functions needed, a test process to measure components adequacy, and documenting functionality and parameters.

#### Process steps for development

Baseline functionality has previously been determined and documented from investigation, experimentation, and tests. This is the functional list of components that provide ample RAP rate and can be labeled as the (RAP list). The RAP list is provided in the section above "*Methods of fixing problems*".

1) Supplier components are investigated for the specific hardware functionality needed. Specifications and documentation distributed by suppliers is investigated for the specific fit to the RAP list of requirements. Process can be manual or automated over the internet,

fax, or US mail. The process can include having the suppliers of hardware components produce the list of hardware that meets the requirements of the RAP list.

2) Verifying standard level of performance available while testing an LVVR application using RAP Rate as an indicator of system standard measurement. Then use industry standard performance measurement tools to isolate areas that need changing or updating focusing on microprocessor, memory, IO subsystem, sound input system while using a LVVR application to determine specific areas of weakness.

3) Document the components and parameters for use in process sheets to be used for manufacture of the standard system. Process sheets include areas for checking if a step was completed and a place for comments at each section. Below is one Process Sheet example to implement manufacture of a standard level LVVR system.

(SEE TABLE 1)

For the manufacture of LVVR systems the process sheets are used as build procedures. These sheets define components to be purchased and packaged, parameter settings for the hardware and software including hardware jumpers, BIOS, operating system parameters, application parameters, and a check sheet for comments back to engineering and repeatability.

(SEE PROCESS 1)

(SEE PROCESS 2)

This process was developed around the concept of achieving a standardized level of Reliable Accuracy and Performance Rate (RAP Rate). The components and functionality needed to achieving an adequate RAP rate are list and explained below, and expected to change over time. However, as the components and technology continue to improve, the standard measurement of RAP rate will remain a valid measurement allowing users to understand what is being purchased or provided. In other words RAP rate is to voice recognition what wattage or power measurement is to the electric industry.

Description of the features in additional detail

A RAP rate of 100/0 defined as 100 % accuracy with zero delay time measure from time of spoken word to displayed text in an application is the ultimate goal. It was determined from research and testing that specific components can affect RAP rate. Additionally where some component may be lacking, another may be more than adequate resulting in a similar RAP rate.

$$RAP\ rate = u + s + q + i$$

The components of the RAP rate equation can be further defined where:

***Quality "q"***

Defined as a compatibility measurement of components and functionality that are well-matched for LVVR.

$$q = CPU\ margin\ \% / (\%\ of\ app\ in\ memory / (\%\ of\ app\ in\ memory - (KB\ Cache / cache\ hit\ rate)/60) - (A/D\ conversion\ time + bus\ throughput\ latency)$$

The equation above indicates aspects of hardware that can be changed to achieve a higher RAP rate focusing on the metric of *Performance.* The quality of the hardware "q" as related to LVVR relates directly to performance. The performance result is measured in time (seconds for current technology). Performance can be a positive or negative value. A positive value indicates delay while a negative value is additional resources that can be used for other task. The "delay" in the performance definition will never be zero. It may not be perceivable to a user, but can be measured by using specific tools designed for these purposes, like the RAP meter.

The following diagram and list are components that can affect RAP rate with today's technology. Fast data access directly relates to the performance in RAP. The order of fastest medium moves out from the microprocessor to disk as shown in the diagram below:

(SEE FIGURE 9)

The following list explains items that are important with these components and list some industry standard methods of measurement. When RAP rate is not acceptable the other methods can be used to isolate the problem areas.
High-speed microprocessor

Microprocessor speeds today are up to 800 MHz+ on average and steadily moving to higher processor speeds. When measuring microprocessor usage while using LVVR applications, results show that microprocessor usage is at 100%. To determine this a combination of a manual process and an automated process is used. One method of measuring CPU usage is by using the performance monitor tools available with an operating system like Microsoft Windows 98.

The goal is to achieve a margin of microprocessor resources left available while dictation to a system is being done. Ideally, with voice recognition a performance in the range of

no noticeable delay from the time the words are spoken to the time the text is displayed in a text editor is a desired metric. If other applications are to be run simultaneously, then an additional margin in performance must be added to avoid affecting RAP rate.

Robust floating point features

A robust floating-point microprocessor is needed due to the intensity of math calculations that are routine for voice recognition applications. Floating point microprocessors may be embedded in a main microprocessor or done separately by the main CPU instruction set or software. Therefore microprocessors that support floating-point in different ways can directly affect RAP rate. Ideally a microprocessor that has a combination of hardware registers, floating point instruction set with features that allow multiple calculations with minimal clock cycles, while supporting access to fast cache memory are desirable. Measurements on floating points can be achieved using industry standard tools or published results in the trade magazines or from the manufacturers.

Large on chip and off chip cache memory

Cache memory is the closest storage medium to the microprocessors doing the work. Typically the memory closest to the main CPU will be the fastest data access The capacity of the cache memory, the percentage of cache hits, and if the cache is embedded in the CPU chip or off chip will make a difference. "*KB Cache / cache hit rate*" work as performance enhancement in the equation and can be measured using embedded OS performance tools of Microsoft Windows.

High-capacity/fast main memory

A large capacity main memory is desired and will affect performance. Enough capacity to allow the LVVR and related applications to execute directly out of memory yields the best performance. Having to go out to disk is a magnitude of time longer and is avoided whenever possible. Testing and measuring results indicate that using a LVVR system can easily use 256 megabytes to prevent disk access. This can be measured using operating system tools like the performance monitor of Microsoft Windows 98, along with other tools available in the computer industry. As memory is reduced a delay resulting in a lower RAP rate will occur. Therefore the equation includes a metric "*% of application in memory*" as add or minus to performance. These values will change over time and technology, but the goal remains the same for LVVR, to execute without disk access.

Sound input device with performance focused in the range of the human voice

Most sound components for PC's focus on output while input is a secondary

consideration. Therefore sound input components can cause performance problems. The physical system interface/bus can also add or subtract to performance. *A/D conversion time + bus throughput latency* time subtracts from the performance and can never be removed from the system. While this delay can be lowered to the level of not perceivable, it will never be reduced to zero. Oscilloscopes are one method of measuring this delay. This measurement is also included in the performance measurement of RAP rate which can be measured through a software tool like a RAP meter.

When the objective of Quality is completed, then integration of the component parameters and reduction in bottle necks are the objective.

***Integration "i"***

*i = # OS parameters + # application parameters – incompatibilities – other task – throughput resistance*

The integration aspect of RAP rate can be affected by software (firmware, operating systems, applications/utilities and parameters). Parameters can enhance or subtract RAP rate from a large vocabulary voice recognition application. As an example, a word processing application with a parameter set to auto correct grammar during dictation may cause sever RAP rate reduction due to resources being shared for real time grammar correction and LVVR.

Starting at the lowest level (BIOS) and working through the OS towards the LVVR application is one method of tuning software for a good RAP rate. Another method would be to reverse the order and start at the LVVR application and work back. Then create a software utility that does the parameter settings automatically based on the known information. Therefore an explanation of the equation above would be to add items that can be modified to enhance LVVR and to subtract items that cannot be removed and must be worked around like incompatibilities. There are not industry standard tools to measure for these types of parameters. At this point RAP rate or an individual component of RAP is the only measurement that sums these conclusions.

Darrell A. Poirier
006-110-300

## LXIV. Large Vocabulary Voice Recognition Handheld Transcriber or "Power Handheld Device" (PHD) that supports that supports VMM and SVMs.

### *Brief History*

The industry standard for dictation is to use handheld tape recorders or memory devices that provide the same functionality as tape recorders. Then some companies provided connections from these handheld devices to desktop type computers allowing the voice to be translated into text through a voice recognition package. These approaches had many problems including:

- No direct feedback while the dictation is taking place
- It was not real time large vocabulary voice recognition
- Training for the voice recognition was cumbersome to accomplish resulting in very poor accuracy and a lot of user frustration. Training also required redundant work of upwards to an hour to be successful.
- Updating the voice parameters and training was typically not possible or very hard to accomplish resulting in the accuracy level not getting better over time.
- Another physical connection to the dictation device was needed to accomplish the translation to text
- Sound quality on playback was frequently poor and could not be understood
- There was little to no control of manipulating the text output until the entire recorded voice was dumped and translated into text

### *Methods of fixing the problems*

After experimenting with handheld transcribers and discovering the problems listed above, several approaches were researched to solve the problems.

One example was simple experimentation using a cable that allowed a microphone to be connected to both the handheld transcriber and a desktop PC and implementing a process of synchronization for training large vocabulary voice recognition on both devices simultaneously.

(SEE FIGURE 7)

This was successful allowing a user to train a hand held transcriber at the same time the desktop system was trained. This method saved the redundant training time. Other experiments were done trying to combine a Voice Model for a desktop VR system with a

hand held transcriber Voice Model in anticipation of having one Voice Model for both desktop system and hand held transcriber. These attempts included:

- Renaming files to match the format of PC Voice Models to handheld transcriber Voice Models with the anticipation of increasing accuracy and remove redundant training.
- Modifying the output .WAV file of a handheld transcriber with the anticipation of a software filter to remove any noise that was not directly related to the Voice Model parameters in effort of increasing accuracy and removing redundant training.
- Experimenting with different microphone types and styles with the anticipation of increasing accuracy connected to the handheld transcriber.

While some of these experiments yielded some success with redundant training and better accuracy, it became clear that the other problems could not be addressed using the current handheld technology. It appeared a better method of accomplishing large vocabulary voice recognition for handheld transcribers would be to package the desktop system hardware into a handheld form factor.

From previous work it was discovered that computer hardware to support fully functional large vocabulary voice recognition for handheld transcribers and these types of applications must include at least the following components to be effective:

- High-speed microprocessor with robust floating point features
- Large on chip and off chip cache
- High-capacity/fast main memory
- Quality sound input device with performance focused in the range of the human voice or signal to be translated
- An operating system specifically configured (tuned) for the application of voice recognition.

Research was done looking for the standard components previously determined in a small size and available off the shelf. These components were found packaged in a small form factor buy a couple of companies that supplied functional components. These smaller components allowed a prototype to be built in 1999 to prove the viability of such a device. This is a picture of the first prototype large vocabulary handheld transcriber using Microsoft Windows 98 and Dragon Professional Voice Dictation application for an example. Other operating systems and applications could have been used like Linux and a public domain voice recognition application. This device solves all the problems listed above.

(SEE FIGURE 8)

The prototype is a fully functioning handheld transcriber focusing on proof of the concepts of form factor, use of VMM via a network drive, the ability to provide direct feedback of speech to text while dictating in the handheld environment, and the ability to use a non-keyboard or mouse voice recognition interface combined with touch screen for

user control. The prototype supports a vocabulary of over 30,000 words. Test results from this proto-type indicate that production models could support large vocabularies including libraries to support medical and legal services. This prototype includes network connection, USB, keyboard and mouse if desired, and connection for 120 volt AC power, and a microphone input jack.

*Theory of operation*

Power on device

After applying power to the device it can be controlled using voice recognition commands and touch screen. When the device becomes ready it automatically is in a mode to select a user or SVM and dictation can start.

Dictate to machine

The device supports a microphone input jack with a microphone on/off switch that can be momentary or left in either mode. The user speaks into a microphone and the voice is translated into a text editor on a handheld screen. What makes this handheld device unique is the amount of words (large vocabulary of greater than 10,000 + words) that can be translated. Translating 5000 words would be considered (SVVR) Small Vocabulary Voice Recognition by these standards.

Save file

The dictated text files can be saved for later editing, importing and exporting, archival or transfer.

Move files over network

The device supports network connection for moving the files and voice models to and from the handheld device.

**LXV. Archive and supply services of voice models**

***Brief History***

Typically when using large vocabulary voice recognition applications, training of the voice recognition software is involved to allow the software to understand the uniqueness of a specific user. This uniqueness is stored in a file, operating system parameter list (like registry entries) or other format that defines specific parameters for a specific user defined by this patent as a Voice Model. The ability to unplug these parameters and data from one machine, put them on some transfer medium like disk, optical disk, floppy, network disk drive, etc. and move them to another machine is defined by this patent as Voice Model Mobility or VMM. Prior to VMM, there was no easy way to move these specific user parameters and data between machines. A voice model that can support several users is defined as a Super Voice Model.

As more people become accustom to a Reliable Accuracy Performance (RAP) rate they will want to enable these features to the future voice recognition technology as it becomes available. At this point in time the Internet is highly popular with industry plans to have internet access in automobiles, appliances, interactive TV, and remote access to in addition to computer access name a few of the applications. Voice Model Mobility and Super Voice Models can be the technology that allows the users to experience the full functionality of these voice recognition Internet technologies.

*The problems*

A problem will arise for people that spend much of their time in a mobile mode of living. Additionally, the different appliances and devices will not likely have common transfer medium technologies with the exception of some type of network connection like the internet. Users will be looking for a method and means to have their personal voice model available for all these devices.

Additionally, technology companies and businesses will likely need voice models that can address many people.

*Methods of fixing problems*

New service providing centrally located archives of Voice Model storage for mobility. This allows a user to have access to a proficient personalized Voice Model regardless of physical location or appliance being used. The voice model can be transferred electronically over wire or wireless transmission. The central voice model can also service multiple locations.

*Theory of operation*

- User creates Voice Model or Super Voice Model is created

- User transfers personalized Voice Model to a location and medium provided by the service provider for fee.
- The location provided by the service provider is accessible from a wide number of device types/appliances like PC's, hand held transcribers, devices for automobiles and house hold use etc.
- When the user moves to a new location, the voice model is available for transfer to the new location.
- Voice Models and Super Voice Models will be available to rent or buy

## LXVI. Voice Mail To Text using large vocabulary voice recognition software and VMM & SVM

### *Brief History*

Voice mail throughout industries and residential use today do not have options for a translation from voice to text using voice recognition software. If this type of output is needed, a human is the best means to get it done. Typically the human would listen to the audio output and put that into a written format. The main problem with translating voice mail to text includes:

- Large vocabularies for voice recognition are needed
- Training or a SVM to allow a person to machine translation is needed for acceptable accuracy
- Voice recognition software packages are not focusing on this need thus not providing needed functionality to enable voice mail systems
- An "ease of use" user interface does not exist to command this type of functionality
- There are no standards for items like RAP rate to ensure a standard level of functionality can be depended upon
- File types and virtual links are not always compatible for voice to text translation between answering machines or voice mail devices to large vocabulary voice recognition software packages.

### *Methods of fixing the problems*

The method of fixing this problem is to combine LVVR applications with systems (hardware and software) that focus on voice mail. Having Voice Models is key to the ability of voice mail to text translation. Voice Models open the door initially to users that have voice models available. Voice mail to text for the masses of people use SVMs as the means. Prior to the invention of using a voice model and VMM or SVM's only limited voice recognition over a telephone (internet or phone system) is achieved. The

advantage of using a voice model in a post processing mode similar to audio data mining is that it allows access to a large vocabulary of words after the voice has been captured.. The main difference from voice mail to text versus data mining is that key words or phases are not being searched for. Instead a full translation from voice to text is the objective.

*Theory of operation*

A person calls on a telephone and leaves an audio message on a storage medium. The message is transferred from an analog format to a digital format like a .WAV file. A specific user's voice model (moved to the machine using VMM or other means) or a super voice model (SVM) that represents a group of users parameters is selected and used in conjunction with a large vocabulary voice recognition application. The voice recognition application translates the audio input into text output. The translation can be completed after the caller hangs up allowing the system resources to focus on accuracy. After the translation has been completed it can be electronically mailed, faxed, or printed for the recipient to read.

Summary of the Inventions

- The concept of separating specific user parameters/information and data from Large Vocabulary Voice Recognition (LVVR) applications .
- Defining the LVVR specific user parameters/information and data as Voice Model
- Ability to transfer the Voice Model to more than a single LVVR application
- Defining the ability to transfer the Voice Model as Voice Model Mobility
- A software application called Voice Model Mobility that runs on an IBM compatible computer
- A software application called Voice Model Mobility that runs on an IBM compatible computer that contains a button that moves a Voice Model
- A software application called Voice Model Mobility that runs on an IBM compatible computer that allows selection of users to transfer
- Abbreviation of Voice Model Mobility to VMM
- Creation of a specific file "VMMinfo.txt"
- Ability to package a Voice Model and create a place for other user specific LVVR information
- Ability to package a Voice Model and create a place for other non-LVVR information
- The translation of a Voice Model from a specific LVVR application to another LVVR application from a different origin or company
- The concept of a Super Voice Model
- Method for creating a Super Voice Model
- The concept of archiving Voice Models
- The concept of creating Voice Models for groups of users

- The concept of synthesizing a Super Voice Model from using known parameters to statistically create the unknown parameters of a user.
- The concept of combining computer voice recognized accuracy and performance into one metric
- Creation of Reliable Accuracy and Performance measurement with regards to LVVR
- Creation of Reliable Accuracy and Performance measurement with regards to LVVR and defining it as RAP Rate.
- Creation of the concept of a specific tool to measure Reliable Accuracy and Performance on LVVR systems
- Creation of an equation that defines RAP Rate
- Creation of an equation that defines performance in RAP Rate
- Creation of an equation that defines accuracy in RAP Rate
- Creation of an equation that defines quality that affects performance in RAP Rate
- Creation of the concept of a specific tool to measure Reliable Accuracy and Performance on LVVR systems labeling it as a RAP Rate Meter or RAP Meter
- The concept of defining RAP Rate syntax as "accuracy %" to "performance time"
- Creating the concept of a certification of a specific RAP Rate measurements
- The process of creating revenues for RAP Rate certified advertising
- Mechanics of building a RAP meter software application
- Creation of the concept of defining components to use for optimal voice recognition using computers
- Creation of the concept of standardized hardware for LVVR applications
- A method of defining specific hardware components for LVVR
- A method of using RAP Rate to measure the standardized hardware
- Creation of a list of component functionality for optimal voice recognition
- A process for development of LVVR systems
- A process for manufacture of LVVR system
- Handheld computer that is IBM compatible
- Handheld computer that has resources equal to that of a desktop computer
- Handheld computer that is a replacement for handheld dictation recorders
- Handheld computer that supports LVVR
- Handheld computer that supports large word vocabularies
- Handheld computer that supports VMM
- Handheld computer that supports direct user feedback/display of words spoken into the LVVR application
- Handheld computer that supports network connection used for VMM
- Handheld computer that does not need LVVR training for voice recognition
- Handheld computer that supports full text editing applications
- The concept of providing services for Voice Models
- The concept of archiving Voice Models
- The concept of supplying Voice Models
- The concept of supplying Voice Models over a network connection
- The concept of selling Voice Models for revenue creation
- The concept of renting Voice Models for revenue creation
- The concept of using LVVR applications to enable voice mail to text products

- The use of VMM to enable LVVR applications in the voice mail to text environment
- The use of SVM to enable LVVR applications in the voice mail to text environment
- The ability to electronically send voice mail in text format fully translated and sent by a machine
- The ability to electronically FAX voice mail in text format fully translated and sent by a machine
- The ability to print voice mail in text format fully translated and printed by a machine

**More Detailed Summary of the Inventions:**

Define what a Voice Model is "concept". Separate voice models from the voice recognition applications. Use VMM to transfer the voice models between systems using the various forms of medium available including network, floppy drives, zip drives, etc. Allow the new voice models to contain other personalized user information.

Inventions:

I. The concept of separating specific user parameters/information and data from Large Vocabulary Voice Recognition (LVVR) applications .

II. Defining the LVVR specific user parameters/information and data as Voice Model

III. Ability to transfer the Voice Model to more than a single LVVR application

IV. Defining the ability to transfer the Voice Model as Voice Model Mobility

V. A software application called Voice Model Mobility that runs on an IBM compatible computer

VI. A software application called Voice Model Mobility that runs on an IBM compatible computer that contains a button that moves a Voice Model

VII. A software application called Voice Model Mobility that runs on an IBM compatible computer that allows selection of users to transfer

VIII. Abbreviation of Voice Model Mobility to VMM

IX. Creation of a specific file "VMMinfo.txt"

X. Ability to package a Voice Model and create a place for other user specific LVVR information

XI. Ability to package a Voice Model and create a place for other non-LVVR information

XII. The translation of a Voice Model from a specific LVVR application to another LVVR application from a different origin or company

Darrell A. Poirier
006-110-300

Define a Super Voice Model. Explain that SVM is a method to address speaker independent voice recognition without training. The method for creating a SVM involves categorizing Voice Models into groups, categorizing people into groups, match the people and SVM. Plug in unknown parameters using a statistical model and create a new type of voice model (SVM).

Inventions:

XIII. The concept of a Super Voice Model
XIV. Method for creating a Super Voice Model
XV. The concept of archiving Voice Models
XVI. The concept of creating Voice Models for groups of users
XVII. The concept of synthesizing a Super Voice Model from using known parameters to statistically create the unknown parameters of a user.

RAP Rate is the ratio measure of accuracy in % with the performance delay in time of displayed text. RAP Rate ratio is measured using a RAP Rate meter. The mechanics of building a meter is function calls to hardware drivers accessed through an interface driven by a user. Calculations are done as the user dictates to a window where incorrect words are flagged and measured, and delay from spoke word to text displayed is measured. Revenues are created through user request of RAP measure certification RAP for LVVR solution providers. Revenues are also provided through sales of VMM where RAP meter is an included option/feature.

Inventions:

XVIII. The concept of combining computer voice recognized accuracy and performance into one metric
XIX. Creation of Reliable Accuracy and Performance measurement with regards to LVVR
XX. Creation of Reliable Accuracy and Performance measurement with regards to LVVR and defining it as RAP Rate.
XXI. Creation of the concept of a specific tool to measure Reliable Accuracy and Performance on LVVR systems
XXII. Creation of an equation that defines RAP Rate
XXIII. Creation of an equation that defines performance in RAP Rate
XXIV. Creation of an equation that defines accuracy in RAP Rate
XXV. Creation of an equation that defines quality that affects performance in RAP Rate
XXVI. Creation of the concept of a specific tool to measure Reliable Accuracy and Performance on LVVR systems labeling it as a RAP Rate Meter or RAP Meter
XXVII. The concept of defining RAP Rate syntax as "accuracy %" to "performance time"
XXVIII. Creating the concept of a certification of a specific RAP Rate measurements
XXIX. The process of creating revenues for RAP Rate certified advertising

XXX. Mechanics of building a RAP meter software application
XXXI. Creation of the concept of defining components to use for optimal voice recognition using computers

Standardized hardware processed defined and measured using RAP Rate as an indicator. The concept is to create and publish standards for optimal RAP rate versus minimum requirements stated by the LVVR applications and hardware providers. The other option is to develop and manufacture LVVR solutions using this process.

Inventions:

XXXII.Creation of the concept of standardized hardware for LVVR applications
XXXIII.A method of defining specific hardware components for LVVR
XXXIV.A method of using RAP Rate to measure the standardized hardware
XXXV.Creation of a list of component functionality for optimal voice recognition
XXXVI.A process for development of LVVR systems
XXXVII.A process for manufacture of LVVR system

Powerful handheld computer as a replacement for handheld dictation recording machines. This device supports LVVR applications, VMM, SVM, and has the physical connections to support VMM. It provides direct feed back via a 640X480 screen. Supports full word processing applications and editing as well as voice control. Others in the industry today run reduced versions of OS with reduced features and capabilities.

Inventions:

XXXVIII.Handheld computer that is IBM compatible
XXXIX.Handheld computer that has resources equal to that of a desktop computer
XL. Handheld computer that is a replacement for handheld dictation recorders
XLI. Handheld computer that supports LVVR
XLII. Handheld computer that supports large word vocabularies
XLIII. Handheld computer that supports VMM
XLIV. Handheld computer that supports direct user feedback/display of words spoken into the LVVR application
XLV. Handheld computer that supports network connection used for VMM
XLVI. Handheld computer that does not need LVVR training for voice recognition
XLVII.Handheld computer that supports full text editing applications

Voice model archive services. These services are web based and provide access to Voice Models and SVMs from multiple locations. The Voice Models can also be sold and bought through this service for use, research, and as components to future applications. This is also the source for SVM creation.

Inventions:

XLVIII.The concept of providing services for Voice Models
XLIX. The concept of archiving Voice Models
L. The concept of supplying Voice Models
LI. The concept of supplying Voice Models over a network connection
LII. The concept of selling Voice Models for revenue creation
LIII. The concept of renting Voice Models for revenue creation

Voice mail to text when done is very limited and requires a human. Using an LVVR application with VMM and SVM enable voice mail to text to be done by machines. For the simple model, the voice mail machine will recognize people that have placed Voice Models on the VMTT machine (corporate use). For the complex model SVMs will provide the ability for the public to use VMTT. VMTT provides voice mail to FAX, e-mail, and print voice messages.

Inventions:

LIV. The concept of using LVVR applications to enable voice mail to text products
LV. The use of VMM to enable LVVR applications in the voice mail to text environment
LVI. The use of SVM to enable LVVR applications in the voice mail to text environment
LVII. The ability to electronically send voice mail in text format fully translated and sent by a machine
LVIII. The ability to electronically FAX voice mail in text format fully translated and sent by a machine
LIX. The ability to print voice mail in text format fully translated and printed by a machine.

While it will be apparent that the illustrated embodiments of the invention herein disclosed are calculated adequately to fulfill the object and advantages primarily stated, it is to be understood that the invention is susceptible to variation, modification, and change within the spirit and scope of the subjoined claims.

The invention having been thus described, what is claimed as new and desire to secure by Letters Patent is: